

THE INVERSE CONJECTURE FOR THE GOWERS NORM OVER FINITE FIELDS VIA THE CORRESPONDENCE PRINCIPLE

TERENCE TAO AND TAMAR ZIEGLER

ABSTRACT. The inverse conjecture for the Gowers norms $U^d(V)$ for finite-dimensional vector spaces V over a finite field \mathbb{F} asserts, roughly speaking, that a bounded function f has large Gowers norm $\|f\|_{U^d(V)}$ if and only if it correlates with a phase polynomial $\phi = e_{\mathbb{F}}(P)$ of degree at most $d - 1$, thus $P : V \rightarrow \mathbb{F}$ is a polynomial of degree at most $d - 1$. In this paper, we develop a variant of the Furstenberg correspondence principle which allows us to establish this conjecture in the large characteristic case $\text{char}(\mathbb{F}) \geq d$ from an ergodic theory counterpart, which was recently established by Bergelson and the authors in [2]. In low characteristic we obtain a partial result, in which the phase polynomial ϕ is allowed to be of some larger degree $C(d)$. The full inverse conjecture remains open in low characteristic; the counterexamples in [13], [15] in this setting can be avoided by a slight reformulation of the conjecture.

1. INTRODUCTION

1.1. The combinatorial inverse conjecture in finite characteristic. Let \mathbb{F} be a finite field of prime order. Throughout this paper, \mathbb{F} will be considered fixed (e.g. $\mathbb{F} = \mathbb{F}_2$ or $\mathbb{F} = \mathbb{F}_3$), and the term “vector space” will be shorthand for “vector space over \mathbb{F} ”, and more generally any linear algebra term (e.g. span, independence, basis, subspace, linear transformation, etc.) will be understood to be over the field \mathbb{F} .

If V is a vector space, $f : V \rightarrow \mathbb{C}$ is a function, and $h \in V$ is a shift, we define the (multiplicative) derivative $\Delta_h f : V \rightarrow \mathbb{C}$ of f by the formula

$$\Delta_h f := (T_h f) \bar{f}$$

where the shift operator T_h with shift h is defined by $T_h f(x) := f(x + h)$. An important special case arises when f takes the form $f = e_{\mathbb{F}}(P)$, where $P : V \rightarrow \mathbb{F}$ is a function, and $e_{\mathbb{F}} : \mathbb{F} \rightarrow \mathbb{C}$ is the standard character $e_{\mathbb{F}}(j) := e^{2\pi i j / |\mathbb{F}|}$ for $j = 0, \dots, |\mathbb{F}| - 1$. In that case we see that $\Delta_h f = e_{\mathbb{F}}(\Delta_h P)$, where $\Delta_h P : V \rightarrow \mathbb{F}$ is the (additive) derivative of P , defined as

$$\Delta_h P = T_h P - P.$$

Given an integer $d \geq 0$, we say that a function $P : V \rightarrow \mathbb{F}$ is a *polynomial of degree at most d* if we have $\Delta_{h_1} \dots \Delta_{h_{d+1}} P = 0$ for all $h_1, \dots, h_{d+1} \in V$, and write $\text{Poly}_d(V)$ for the set of all polynomials on V of degree at most d ; thus for instance $\text{Poly}_0(V)$ is

The first author is supported by a grant from the MacArthur Foundation, and by NSF grant CCF-0649473. The second author is supported by ISF grant 557/08, by a Landau fellowship - supported by the Taub foundations, and by an Alon fellowship.

the set of constants, $\text{Poly}_1(V)$ is the set of linear polynomials on V , $\text{Poly}_2(V)$ is the set of quadratic polynomials, and so forth. It is easy to see that $\text{Poly}_d(V)$ is a vector space, and if $V = \mathbb{F}^n = \{(x_1, \dots, x_n) : x_1, \dots, x_n \in \mathbb{F}\}$ is the standard n -dimensional vector space, then $\text{Poly}_d(V)$ has the monomials $x_1^{i_1} \dots x_n^{i_n}$ for $0 \leq i_1, \dots, i_n < |\mathbb{F}|$ and $i_1 + \dots + i_n \leq d$ as a basis¹.

We shall say that a function $f : V \rightarrow \mathbb{C}$ is a *phase polynomial of degree at most d* if all $(d+1)^{\text{th}}$ multiplicative derivatives $\Delta_{h_1} \dots \Delta_{h_{d+1}} f$ are identically 1, and write $\mathcal{P}_d(V)$ for the space of all phase polynomials of degree at most d . We have the following equivalence between polynomials and phase polynomials in the high characteristic case:

Lemma 1.2 (Phase polynomials are exponentials of polynomials). *Suppose that $0 \leq d < \text{char}(\mathbb{F})$, and $f : V \rightarrow \mathbb{C}$. Then the following are equivalent:*

- (i) $f \in \mathcal{P}_d(V)$.
- (ii) $f = e^{2\pi i \theta} e_{\mathbb{F}}(P)$ for some $\theta \in \mathbb{R}/\mathbb{Z}$ and $P \in \text{Poly}_d(V)$.

Proof. See [2, Lemma D.5]. □

Remark 1.3. The lemma fails in the low characteristic case $d \geq \text{char}(\mathbb{F})$; consider for instance the function $f : \mathbb{F}_2 \rightarrow \mathbb{C}$ defined by $f(1) := i$ and $f(0) := 1$. This function lies in $\mathcal{P}_2(\mathbb{F}_2)$ but does not arise from a polynomial in $\text{Poly}_2(\mathbb{F}_2)$.

Definition 1.4 (Expectation notation). If A is a finite non-empty set and $f : A \rightarrow \mathbb{C}$ is a function, we write $|A|$ for the cardinality of A , and $\mathbb{E}_A f$, $\int_A f$, or $\mathbb{E}_{x \in A} f(x)$ for the average $\frac{1}{|A|} \sum_{x \in A} f(x)$.

Definition 1.5 (Gowers uniformity norm). [6], [7] Let V be a finite vector space, let $f : V \rightarrow \mathbb{C}$ be a function, and let $d \geq 1$ be an integer. We then define the *Gowers norm* $\|f\|_{U^d(V)}$ of f to be the quantity

$$\|f\|_{U^d(V)} := |\mathbb{E}_{h_1, \dots, h_d} \int_V \Delta_{h_1} \dots \Delta_{h_d} f|^{1/2^d},$$

thus $\|f\|_{U^{d+1}(V)}$ measures the average bias in d^{th} multiplicative derivatives of f . We also define the *weak Gowers norm* $\|f\|_{u^d(V)}$ of f to be the quantity

$$\|f\|_{u^d(V)} := \sup_{\phi \in \mathcal{P}_{d-1}(V)} \left| \int_V f \bar{\phi} \right|, \tag{1.1}$$

thus $\|f\|_{u^d(V)}$ measures the extent to which f can correlate with a phase polynomial of degree at most $d-1$.

Remark 1.6. It can in fact be shown that the Gowers and weak Gowers norm are in fact norms for $d \geq 2$ (and seminorms for $d = 1$), see e.g. [7] or [21]. Further discussion of these two norms can be found in [10]. In view of Lemma 1.2, in the high characteristic case $\text{char}(\mathbb{F}) \geq d$ one can replace the phase polynomial $\phi \in \mathcal{P}_{d-1}(V)$ in (1.1) by the exponential $e_{\mathbb{F}}(P)$ of a polynomial $P \in \text{Poly}_{d-1}(V)$. However, this is not the case in low

¹The restriction $i_1, \dots, i_n < |\mathbb{F}|$ arises of course from the identity $x^{|\mathbb{F}|} = x$ for all $x \in \mathbb{F}$.

characteristic. For instance, let $\mathbb{F} = \mathbb{F}_2$, $V = \mathbb{F}_2^n$, and consider the symmetric function $S_4 : V \rightarrow \mathbb{F}_2$ defined by

$$S_4(x_1, \dots, x_n) := \sum_{1 \leq i < j < k < l \leq n} x_i x_j x_k x_l.$$

Then the function $f := (-1)^{S_4}$ has low correlation with any exponential $e_{\mathbb{F}}(P) = (-1)^P$ of a cubic polynomial $P \in \text{Poly}_3(V)$ in the sense that $\mathbb{E}_{x \in V} f e_{\mathbb{F}}(-P) = o_{n \rightarrow \infty}(1)$ (see [15], [13]); on the other hand, it is not hard to verify that the function

$$g(x_1, \dots, x_n) := e^{2\pi i |x|/8},$$

where $|x|$ denotes the number of indices $1 \leq j \leq n$ for which $x_j = 1$, lies in $\mathcal{P}_3(V)$ and has a large inner product with f ; indeed, since $f(x) = +1$ when $|x| = 0, 1, 2, 3 \pmod{8}$ and -1 otherwise, we easily check that

$$\begin{aligned} \mathbb{E}_{x \in V} f \bar{g} &= \frac{1}{8} (1 + e^{-\frac{2\pi i}{8}} + e^{-\frac{4\pi i}{8}} + e^{-\frac{6\pi i}{8}} - e^{-\frac{8\pi i}{8}} - e^{-\frac{10\pi i}{8}} - e^{-\frac{12\pi i}{8}} - e^{-\frac{14\pi i}{8}}) + o_{n \rightarrow \infty}(1) \\ &= \frac{1 - i - \sqrt{2}i}{4} + o_{n \rightarrow \infty}(1). \end{aligned}$$

In particular, we see that $\|(-1)^{S_4}\|_{u^4(V)}$ is bounded from below by a positive absolute constant for large n .

Let $\mathcal{D} := \{z \in \mathbb{C} : |z| \leq 1\}$ be the compact unit disk. This paper is concerned with the following conjecture:

Conjecture 1.7 (Inverse Conjecture for the Gowers norm). *Let \mathbb{F} be a finite field and let $d \geq 1$ be an integer. Then for every $\delta > 0$ there exists $\varepsilon > 0$ such that $\|f\|_{u^d(V)} \geq \varepsilon$ for every finite vector space V and every function $f : V \rightarrow \mathcal{D}$ such that $\|f\|_{U^d(V)} \geq \delta$.*

Remark 1.8. This result is trivial for $d = 1$, and follows easily from Plancherel's theorem for $d = 2$. The result was established for $d = 3$ in [10] (for odd characteristic) and [17] (for even characteristic), and a formulation of Theorem 1.9 was then conjectured in both papers, in which the phase polynomials were constrained to be $\text{char}(\mathbb{F})^{\text{th}}$ roots of unity. This formulation of the conjecture turned out to fail in the low characteristic regime $\text{char}(\mathbb{F}) + 1 < d$ (see [13], [15]); however, the counterexamples given there do not rule out the conjecture as formulated above in this case, basically because of the discussion in Remark 1.6.

The case when δ was sufficiently close to 1 (depending on d) was treated in [1], while the case when $\text{char}(\mathbb{F})$ is large compared to d and δ was established in [18]. In [13], Theorem 1.9 was also established in the case when f was a phase polynomial of degree less than $\text{char}(\mathbb{F})$. These results have applications to solving linear systems of equations (and in particular, in finding arithmetic progressions) in subsets of vector spaces [12], [8] and also to polynomiality testing [17], [3]. Conjecture 1.7 is also the finite field analogue of a corresponding inverse conjecture for the Gowers norm in cyclic groups $\mathbb{Z}/N\mathbb{Z}$, which is of importance in solving linear systems of equations in sets of integers such as the primes; see [11], [4] for further discussion.

The main result of this paper is to establish this conjecture in the high characteristic case.

Theorem 1.9 (Inverse Conjecture for the Gowers norm in high characteristic). *Conjecture 1.7 holds whenever $\text{char}(\mathbb{F}) \geq d$.*

In the low characteristic case we have a partial result:

Theorem 1.10 (Partial inverse Conjecture for the Gowers norm). *Let \mathbb{F} be a finite field and let $d \geq 1$ be an integer. Then for every $\delta > 0$ there exists $\varepsilon > 0$ such that $\|f\|_{u^k(V)} \geq \varepsilon$ for every finite vector space V and every function $f : V \rightarrow \mathcal{D}$ such that $\|f\|_{U^d(V)} \geq \delta$, where $k = C(d)$ depends only on d .*

Remark 1.11. One could in principle make the quantity $k = C(d)$ in Theorem 1.10 explicit, but this would require analyzing the arguments in [2] in careful detail. One should however be able to obtain reasonable values of k for small d (e.g. $d = 4$).

The proofs of Theorems 1.9, 1.10 rely on four additional ingredients:

- An ergodic inverse theorem for the Gowers norm for \mathbb{F}^ω -systems (Theorems 1.19, 1.20), established in [2];
- The Furstenberg correspondence principle[5], combined with the random averaging trick of Varnavides[22];
- A statistical sampling lemma (Proposition 3.13); and
- Local testability of phase polynomials (Lemma 4.5), essentially established in [1].

Of these ingredients, the ergodic inverse theorem is the most crucial, and we now pause to describe it in detail.

1.12. The ergodic inverse conjecture in finite characteristic. Let $\mathbb{F}^\omega := \bigcup_{n=0}^\infty \mathbb{F}^n$ be the inverse limit of the finite-dimensional vector spaces \mathbb{F}^n , where each \mathbb{F}^n is included in the next space \mathbb{F}^{n+1} in the obvious manner; equivalently, \mathbb{F}^ω is the space of sequences $(x_i)_{i=1}^\infty$ with $x_i \in \mathbb{F}$, and all but finitely many of the x_i non-zero. This is a countably infinite vector space over \mathbb{F} .

Definition 1.13 (\mathbb{F}^ω -system). A \mathbb{F}^ω -system is a quadruplet $\mathbf{X} = (X, \mathcal{B}, \mu, (T_g)_{g \in \mathbb{F}^\omega})$, where (X, \mathcal{B}, μ) is a probability space, and $T : h \mapsto T_h$ is a measure-preserving action of \mathbb{F}^ω on X , thus for each $h \in \mathbb{F}^\omega$, $T_h : X \rightarrow X$ is a measure-preserving bijection such that $T_h \circ T_k = T_{h+k}$ for all $h, k \in \mathbb{F}^\omega$. Given any measurable $\phi : X \rightarrow \mathbb{C}$ and $h \in \mathbb{F}^\omega$, we define $T_h \phi : X \rightarrow \mathbb{C}$ to be the function $T_h \phi := \phi \circ T_h$, and $\Delta_h \phi : X \rightarrow \mathbb{C}$ to be the function $\Delta_h \phi := T_h \phi \cdot \overline{\phi}$. We also define the inner product $\langle f, g \rangle := \int_X f \overline{g} \, d\mu$ for all $f, g \in L^2(\mathbf{X})$, where the Lebesgue spaces $L^p(\mathbf{X}) = L^p(X, \mathcal{B}, \mu)$ are defined in the usual manner. We say that the system is *ergodic* if the only \mathbb{F}^ω -invariant functions on $L^2(\mathbf{X})$ are the constants.

Definition 1.14 (Phase polynomial). Let $\mathbf{X} = (X, \mathcal{B}, \mu, (T_g)_{g \in \mathbb{F}^\omega})$ be an \mathbb{F}^ω -system, and let $d \geq 0$. We say that a function $\phi \in L^\infty(\mathbf{X})$ is a *phase polynomial of degree at*

most d if we have $\Delta_{h_1} \dots \Delta_{h_{d+1}} \phi = 1$ μ -a.e. for all $h_1, \dots, h_{d+1} \in \mathbb{F}^\omega$. We let $\mathcal{P}_d(\mathbf{X})$ denote the space of all phase polynomials.

Remark 1.15. By setting $h_1 = \dots = h_{d+1} = 0$ we see that every phase polynomial $\phi \in \mathcal{P}_d(\mathbf{X})$ has unit magnitude: $|\phi| = 1$ μ -a.e..

Definition 1.16 (Gowers-Host-Kra seminorms). [14] Let $\mathbf{X} = (X, \mathcal{B}, \mu, (T_g)_{g \in \mathbb{F}^\omega})$ be a \mathbb{F}^ω -system, and let $\phi \in L^\infty(\mathbf{X})$. We define the *Gowers-Host-Kra seminorms* $\|\phi\|_{U^d(\mathbf{X})}$ for $d \geq 1$ recursively as follows:

- If $d = 1$, then $\|\phi\|_{U^1(\mathbf{X})} := \limsup_{n \rightarrow \infty} \left(\|\mathbb{E}_{h \in \mathbb{F}^n} T_h \phi\|_{L^2(X, \mu)}^2 \right)^{1/2}$;
- If $d > 1$, then $\|\phi\|_{U^d(\mathbf{X})} := \limsup_{n \rightarrow \infty} \left(\|\Delta_h \phi\|_{U^{d-1}(X, \mu, T)}^{2^{d-1}} \right)^{1/2^d}$.

We also define the *weak Gowers-Host-Kra seminorm* $\|\phi\|_{u^d(\mathbf{X})}$ as

$$\|\phi\|_{u^d(\mathbf{X})} := \sup_{\psi \in \mathcal{P}_{d-1}(\mathbf{X})} |\langle \phi, \psi \rangle|.$$

Example 1.17. If $\phi \in \mathcal{P}_{d-1}(\mathbf{X})$ is a phase polynomial of degree at most $d - 1$, then $\|\phi\|_{U^d(\mathbf{X})} = \|\phi\|_{u^d(\mathbf{X})} = 1$.

Remark 1.18. One can use the ergodic theorem to show that the limits here in fact converge, but we will not need this. The U^d are indeed seminorms, but we will not need this either.

In [2, Corollaries 1.26, 1.27], the following ergodic theory analogues of Theorems 1.9, 1.10 was shown:

Theorem 1.19 (Inverse Conjecture for the Gowers-Host-Kra seminorm for high characteristic). *Let $\mathbf{X} = (X, \mathcal{B}, \mu, (T_g)_{g \in \mathbb{F}^\omega})$ be an ergodic \mathbb{F}^ω -system, let $\text{char}(\mathbb{F}) \geq d \geq 1$, and let $\phi \in L^\infty(\mathbf{X})$ be such that $\|\phi\|_{U^d(\mathbf{X})} > 0$. Then $\|\phi\|_{u^d(\mathbf{X})} > 0$.*

Theorem 1.20 (Partial Inverse Conjecture for the Gowers-Host-Kra seminorm for general characteristic). *Let $\mathbf{X} = (X, \mathcal{B}, \mu, (T_g)_{g \in \mathbb{F}^\omega})$ be an ergodic \mathbb{F}^ω -system, let $d \geq 1$, and let $\phi \in L^\infty(\mathbf{X})$ be such that $\|\phi\|_{U^d(\mathbf{X})} > 0$. Then $\|\phi\|_{u^k(\mathbf{X})} > 0$ for some $k = C(d)$ depending only on d .*

Remark 1.21. The “if” part of this theorem follows easily from van der Corput’s lemma; the important part of the theorem for us is the “only if” part. These results can be viewed as a finite field analogue of the results in [14] in high characteristic (and a partial analogue in the low characteristic case), and indeed draws heavily on the tools developed in that paper; see [2] for further discussion. It is quite possible that k can in fact be taken to equal d in Theorem 1.20 (or equivalently, that the condition $\text{char}(\mathbb{F}) \geq d$ can be dropped in Theorem 1.19); this would imply Conjecture 1.7 in full generality.

We will use Theorem 1.20 as a “black box”, and it will be the primary ingredient in our proof of Theorem 1.10, in much the same way that the Furstenberg recurrence theorem is the primary ingredient in Furstenberg’s proof of Szemerédi’s theorem in [5]. Theorem 1.19 plays a similar role for Theorem 1.9.

As with any other argument using a Furstenberg-type correspondence principle, our bounds are *ineffective*, in that we do not obtain an explicit value of ε in terms of d and δ . In principle, one could finitise the arguments in [2] (in the spirit of [19]) to obtain such an explicit value, but this would be extremely tedious (and not entirely straightforward), and would lead to an extremely poor dependence (such as iterated tower-exponential or worse). We will not pursue this matter here.

1.22. Acknowledgments. The first author is supported by a grant from the MacArthur Foundation, and by NSF grant CCF-0649473. The second author is supported by ISF grant 557/08, by a Landau fellowship - supported by the Taub foundations, and by an Alon fellowship. The authors are also greatly indebted to Ben Green for helpful conversations, and Vitaly Bergelson for encouragement.

2. NOTATION

We will rely heavily on asymptotic notation. Given any parameters x_1, \dots, x_k , we use $O_{x_1, \dots, x_k}(X)$ to denote any quantity bounded in magnitude by $C_{x_1, \dots, x_k}X$ for some finite quantity C_{x_1, \dots, x_k} depending only on x_1, \dots, x_k . We also write $Y \ll_{x_1, \dots, x_k} X$ or $X \gg_{x_1, \dots, x_k} Y$ for $Y = O_{x_1, \dots, x_k}(X)$. Furthermore, given an asymptotic parameter n that can go to infinity, we use $o_{n \rightarrow \infty; x_1, \dots, x_k}(X)$ to denote any quantity bounded in magnitude by $c_{x_1, \dots, x_k}(n)X$, where $c_{x_1, \dots, x_k}(n)$ is a quantity which goes to zero as $n \rightarrow \infty$ for fixed x_1, \dots, x_k . Thus for instance, if $r_2 > r_1 > 1$, then $\frac{\exp(r_1)}{\log r_2} = o_{r_2 \rightarrow \infty; r_1}(1)$.

3. STATISTICAL SAMPLING

It is well known that the “global average” $\mathbb{E}_{h \in V} f(h)$ of a bounded function $f : V \rightarrow \mathcal{D}$ can be accurately estimated (with high probability) by randomly selecting a number of points $x_1, \dots, x_N \in V$ and computing the empirical Monte Carlo average (or “local average”) $\mathbb{E}_{1 \leq n \leq N} f(x_n)$. Indeed, it is not hard to show (by the second moment method) that with probability $o_{N \rightarrow \infty}(1)$, one has

$$\mathbb{E}_{1 \leq n \leq N} f(x_n) = \mathbb{E}_{h \in V} f(h) + o_{N \rightarrow \infty}(1).$$

The point here is that the error term is uniform in the choice of f and V .

We now record some variants of this standard “random local averages approximate global averages” fact, in which we perform more exotic empirical averages. We begin with averages along random subspaces of V .

Lemma 3.1 (Random sampling for integrals). *Let V be a finite-dimensional vector space, and let $f : V \rightarrow \mathcal{D}$ be a function. Let $v_1, \dots, v_m \in V$ be chosen independently at random. Then with probability $1 - o_{m \rightarrow \infty}(1)$, we have*

$$\mathbb{E}_{\vec{a} \in \mathbb{F}^m} f(\vec{a} \cdot \vec{v}) = \mathbb{E}_{h \in V} f(h) + o_{m \rightarrow \infty}(1)$$

where $\vec{v} := (v_1, \dots, v_m)$, and $\vec{a} \cdot \vec{v} := a_1 v_1 + \dots + a_m v_m$.

Remark 3.2. One can easily make the $o_{m \rightarrow \infty}(1)$ terms more explicit, but we will not need to do so here.

Proof. We use the second moment method. Note that

$$\mathbb{E}\mathbb{E}_{\vec{a} \in \mathbb{F}^m} f(\vec{a} \cdot \vec{v}) = \mathbb{E}_{h \in V} f(h) + o_{m \rightarrow \infty}(1)$$

(the $o_{m \rightarrow \infty}(1)$ error arising from the $a = 0$ contribution) so by Chebyshev's inequality it suffices to show that

$$\mathbb{E}|\mathbb{E}_{\vec{a} \in \mathbb{F}^m} f(\vec{a} \cdot \vec{v})|^2 = |\mathbb{E}_{h \in V} f(h)|^2 + o_{m \rightarrow \infty}(1).$$

The left-hand side can be rearranged as

$$\mathbb{E}_{\vec{a}, \vec{b} \in \mathbb{F}^m} \mathbb{E} f(\vec{a} \cdot \vec{v}) \bar{f}(\vec{b} \cdot \vec{v}).$$

It is easy to see that the inner expectation is $|\mathbb{E}_{h \in V} f(h)|^2$ unless $\vec{a} = c\vec{b}$, for some $c \in \mathbb{F}$ in which case it is $O(1)$. The claim follows. \square

In the above lemma, f was deterministic and thus independent of the v_i . But we can easily extend the result to the case where f depends on a bounded number of the v_i :

Corollary 3.3 (Random sampling for integrals, II). *Let V be a finite-dimensional vector space, let $m \geq m_0 \geq 0$, let $v_1, \dots, v_m \in V$ be chosen independently at random, and let $f_{v_1, \dots, v_{m_0}} : V \rightarrow \mathcal{D}$ be a function that depends on v_1, \dots, v_{m_0} but is independent of v_{m_0+1}, \dots, v_m . Then with probability $1 - o_{m \rightarrow \infty; m_0}(1)$, we have*

$$\mathbb{E}_{\vec{a} \in \mathbb{F}^m} f_{v_1, \dots, v_{m_0}}(\vec{a} \cdot \vec{v}) = \mathbb{E}_{h \in V} f_{v_1, \dots, v_{m_0}}(h) + o_{m \rightarrow \infty; m_0}(1).$$

Proof. We write $\vec{a} = (\vec{a}_0, \vec{a}_1) \in \mathbb{F}^{m_0} \times \mathbb{F}^{m-m_0}$ and $\vec{v} = (\vec{v}_0, \vec{v}_1) \in V^{m_0} \times V^{m-m_0}$. If we condition $\vec{v}_0 = (v_1, \dots, v_{m_0})$ to be fixed, we see from applying Lemma 3.1 to the remaining random vectors \vec{v}_1 that for fixed \vec{a}_0 , we have

$$\mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{m-m_0}} f_{v_1, \dots, v_{m_0}}(\vec{a} \cdot \vec{v}) = \mathbb{E}_{h \in V} f_{v_1, \dots, v_{m_0}}(\vec{a}_0 \cdot \vec{v}_0 + h) + o_{m-m_0 \rightarrow \infty}(1)$$

with probability $1 - o_{m-m_0 \rightarrow \infty}(1)$ conditioning on \vec{v}_0 ; integrating this we see that the same is true without the conditioning. We can shift h by $\vec{a}_0 \cdot \vec{v}_0$, move the h average onto the other side, and take expectations to conclude that

$$\mathbb{E}|\mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{m-m_0}} f_{v_1, \dots, v_{m_0}}(\vec{a} \cdot \vec{v}) - \mathbb{E}_{h \in V} f_{v_1, \dots, v_{m_0}}(h)| = o_{m-m_0 \rightarrow \infty}(1)$$

for each \vec{a}_0 ; averaging over \vec{a}_0 by the triangle inequality we obtain the claim. \square

Remark 3.4. It is with this corollary that we are implicitly exploiting the highly transitive nature of the symmetry group $GL(V)$ available to us. In the setting of the cyclic group $\mathbb{Z}/N\mathbb{Z}$, the analogue of Lemma 3.1 is still true, namely that one can approximate a global average $\int_{\mathbb{Z}/N\mathbb{Z}} f$ by a local average on random arithmetic progressions of medium length, but this approximation no longer holds if f is allowed to depend on the first few values of that progression, since this of course determines the rest of the progression; this is related to the fact that (for N prime, say), the affine group of $\mathbb{Z}/N\mathbb{Z}$ (which is analogous to $GL(V)$) is 2-transitive but no stronger. In contrast, in the finite field setting, a small subspace of a medium-dimensional subspace does not determine the whole subspace.

We will need to generalise these results further by considering more exotic averages along cubes. A typical result we will need can be stated informally as

$$\begin{aligned} & \mathbb{E}_{\vec{a}_2 \in \mathbb{F}^{m_2}} \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{m_1}} \int_V f(T_{\vec{a}_1 \cdot \vec{v}_1} \bar{f})(T_{\vec{a}_2 \cdot \vec{v}_2} \bar{f})(T_{\vec{a}_1 \cdot \vec{v}_1 + \vec{a}_2 \cdot \vec{v}_2} f) \\ & \approx \mathbb{E}_{h_1, h_2 \in V} \int_V f(T_{h_1} \bar{f})(T_{h_2} \bar{f}) T_{h_1 + h_2} f \end{aligned} \quad (3.1)$$

when m_1 is large, m_2 is large compared with m_1 , and \vec{v} is random (see Lemma 3.9 for the formal version of this type of estimate). Such results follow (heuristically, at least), by iterating the previous results. For instance, from Corollary 3.3 we heuristically have

$$\begin{aligned} & \mathbb{E}_{\vec{a}_2 \in \mathbb{F}^{m_2}} \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{m_1}} \int_V f(T_{\vec{a}_1 \cdot \vec{v}_1} \bar{f})(T_{\vec{a}_2 \cdot \vec{v}_2} \bar{f})(T_{\vec{a}_1 \cdot \vec{v}_1 + \vec{a}_2 \cdot \vec{v}_2} f)(x) \\ & \approx \mathbb{E}_{h_2 \in V} \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{m_1}} \int_V f(T_{\vec{a}_1 \cdot \vec{v}_1} \bar{f})(T_{h_2} \bar{f})(T_{\vec{a}_1 \cdot \vec{v}_1 + h_2} f) \end{aligned}$$

when m_2 is large compared to m_1 and then interchanging the expectations and applying Lemma 3.1 heuristically yields

$$\begin{aligned} & \mathbb{E}_{h_2 \in V} \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{m_1}} \int_V f(T_{\vec{a}_1 \cdot \vec{v}_1} \bar{f})(T_{h_2} \bar{f})(T_{\vec{a}_1 \cdot \vec{v}_1 + h_2} f) \\ & \approx \mathbb{E}_{h_1 \in V} \mathbb{E}_{h_2 \in V} \int_V f(T_{h_1} \bar{f})(T_{h_2} \bar{f})(T_{h_1 + h_2} f) \end{aligned}$$

when m_1 is large, thus giving (3.1).

We will formalise the precise statement along these lines that we need later in this section. We begin with some key definitions.

Definition 3.5 (Lipschitz norm). If $G : \mathcal{D}^n \rightarrow \mathbb{C}$ is a function on a polydisk \mathcal{D}^n , we define the *Lipschitz norm* $\|G\|_{\text{Lip}}$ of G to be the quantity

$$\|G\|_{\text{Lip}} := \sup_{z \in \mathcal{D}^n} |G(z)| + \sup_{z, w \in \mathcal{D}^n : z \neq w} \frac{|G(z) - G(w)|}{d(z, w)}$$

where we use the metric

$$d((z_1, \dots, z_n), (w_1, \dots, w_n)) := |z_1 - w_1| + \dots + |z_n - w_n|.$$

Definition 3.6 (Accurate sampling sequence). Let $k \geq 1$, let V be a finite-dimensional vector space, let $f : V \rightarrow \mathcal{D}$ be a bounded function, and let

$$0 = H_0 < H_1 < H_2 < H_3 < \dots$$

be a sequence of integers (or “scales”). We define an *accurate sampling sequence* for f of degree k and at scales H_1, H_2, \dots to be an infinite sequence of vectors

$$v_1, v_2, v_3, \dots \in V$$

such that for every sequence

$$0 \leq r_0 < r_1 < \dots < r_k$$

of scales and every Lipschitz function $G : \mathcal{D}^{\{0,1\}^k \times \mathbb{F}^{H_{r_0}}} \rightarrow \mathbb{C}$, we have

$$\int_V |G_{f, r_0, r_1, \dots, r_k} - G_{f, r_0}| \leq \frac{\|G\|_{\text{Lip}}}{r_1} \quad (3.2)$$

where

$$G_{f,r_0,r_1,\dots,r_k}(x) := \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{H_{r_1}}, \dots, \vec{a}_k \in \mathbb{F}^{H_{r_k}}} G((f(x + \omega \cdot \mathbf{u} + \vec{b} \cdot \vec{v}_0))_{\omega \in \{0,1\}^k, \vec{b} \in \mathbb{F}^{H_{r_0}}}),$$

where

$$\mathbf{u} = (\vec{a}_1 \cdot \vec{v}_1, \dots, \vec{a}_k \cdot \vec{v}_k); \quad \vec{v}_j = (v_1, \dots, v_{H_{r_j}}), \quad j = 0, \dots, k,$$

and

$$G_{f,r_0}(x) := \mathbb{E}_{h_1 \in V, \dots, h_k \in V} G((f(x + \omega \cdot \mathbf{h} + \vec{b} \cdot \vec{v}_0))_{\omega \in \{0,1\}^k, \vec{b} \in \mathbb{F}^{H_{r_0}}}),$$

where $\mathbf{h} = (h_1, \dots, h_k)$.

Remark 3.7. The denominator r_1 in (3.2) could be replaced by any other fixed function of r_1 that went to infinity as $r_1 \rightarrow \infty$ if desired here.

Remark 3.8. We make the trivial but useful remark that an accurate sampling sequence of degree k is also an accurate sampling sequence of degree k' for any $1 \leq k' \leq k$. Indeed, to verify (3.2) for a function $G' : \mathcal{D}^{\{0,1\}^{k'} \times \mathbb{F}^{H_{r_0}}} \rightarrow \mathcal{D}$ and some scales $r_{k'} > \dots > r_0 \geq 0$, one simply adds some dummy scales $r_{k'+1}, \dots, r_k$ above $r_{k'}$ and extends G' to a function $G : \mathcal{D}^{\{0,1\}^k \times \mathbb{F}^{H_{r_0}}} \rightarrow \mathcal{D}$ by composing with the obvious restriction map from $\mathcal{D}^{\{0,1\}^k \times \mathbb{F}^{H_{r_0}}}$ to $\mathcal{D}^{\{0,1\}^{k'} \times \mathbb{F}^{H_{r_0}}}$.

Roughly speaking, an accurate sampling sequence will allow us to estimate all the global averages that we need for the combinatorial inverse conjecture for the Gowers norm by local averages which are suitable for lifting to the ergodic setting via the correspondence principle. We illustrate the use of such sequences by describing the three special cases of (3.2) that we will actually need in our arguments.

Lemma 3.9 (Global Gowers norm can be approximated by local Gowers norm). *Let $d \geq 1$, let V be a finite-dimensional vector space, let $f : V \rightarrow \mathcal{D}$ be a bounded function, and let $v_1, v_2, \dots \in V$ be an accurate sampling sequence for f of degree d and at scales H_1, H_2, \dots . Then for every sequence of scales*

$$0 < r_1 < r_2 < \dots < r_d$$

we have

$$\mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{H_{r_1}}, \dots, \vec{a}_d \in \mathbb{F}^{H_{r_d}}} \int_V \Delta_{\vec{a}_1 \cdot \vec{v}_{r_1}} \dots \Delta_{\vec{a}_d \cdot \vec{v}_{r_d}} f = \|f\|_{U^d(V)}^{2^d} + o_{r_1 \rightarrow \infty; d}(1).$$

Remark 3.10. As with all other estimates in this section, the point is that the error term is uniform over all choices of f and V . Note that the $d = 2$ case of this lemma is a formalisation of (3.1).

Proof. We apply (3.2) with $r_0 = 0$, and $G : \mathcal{D}^{\{0,1\}^d} \rightarrow \mathbb{C}$ being the function

$$G((z(\omega))_{\omega \in \{0,1\}^d}) := \prod_{\omega \in \{0,1\}^d} \mathcal{C}^{\omega_1 + \dots + \omega_d} z(\omega)$$

where $\mathcal{C} : z \mapsto \bar{z}$ is the complex conjugation operator. A routine computation gives the identities

$$G_{f,0,r_1,\dots,r_d}(x) = \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{H_{r_1}}, \dots, \vec{a}_d \in \mathbb{F}^{H_{r_d}}} \Delta_{\vec{a}_1 \cdot \vec{v}_{r_1}} \dots \Delta_{\vec{a}_d \cdot \vec{v}_{r_d}} f$$

and

$$\int_V G_{f,0} = \|f\|_{U^d(V)}^{2^d}.$$

Also, it is easy to see that the Lipschitz norm $\|G\|_{\text{Lip}}$ is $O_d(1)$. The claim now follows immediately from (3.2) and the triangle inequality. \square

Lemma 3.11 (Global averages can be approximated by local averages). *Let V be a finite-dimensional vector space, let $f : V \rightarrow \mathcal{D}$ be a bounded function, and let $v_1, v_2, \dots \in V$ be an accurate sampling sequence for f of degree 1 and at scales H_1, H_2, \dots . Then for every finite sequence $\vec{b}_1, \dots, \vec{b}_m \in \mathbb{F}^\omega$ and every continuous function $F : \mathcal{D}^m \rightarrow \mathbb{C}$, we have*

$$\int_V |\mathbb{E}_{\vec{a} \in \mathbb{F}^{H_r}} T_{\vec{a} \cdot \vec{v}} g - \int_V g| = o_{r \rightarrow \infty; F, m, \vec{b}_1, \dots, \vec{b}_m}(1)$$

where $g : V \rightarrow \mathbb{C}$ is the function

$$g(x) := F(T_{\vec{b}_1 \cdot \vec{v}} f(x), \dots, T_{\vec{b}_m \cdot \vec{v}} f(x)). \quad (3.3)$$

Proof. By approximating the continuous function F uniformly by a Lipschitz function, we may assume that F is Lipschitz. By adding dummy vectors to the collection $\vec{b}_1, \dots, \vec{b}_m$ if necessary, we may assume that $\{\vec{b}_1, \dots, \vec{b}_m\} = \mathbb{F}^{H_{r_0}}$ for some $r_0 > 0$ depending on $\vec{b}_1, \dots, \vec{b}_m$, thus F is now a Lipschitz function from $\mathcal{D}^{\mathbb{F}^{H_{r_0}}}$ to \mathbb{C} .

Note that to prove the claim we may without loss of generality restrict to the regime $r > r_0$. We now apply (3.2) with $G : \mathcal{D}^{\{0,1\} \times \mathbb{F}^{H_{r_0}}} \rightarrow \mathbb{C}$ being the function

$$G\left((z(\omega, \vec{b}))_{\omega \in \{0,1\}, \vec{b} \in \mathbb{F}^{H_{r_0}}}\right) := F\left((z(1, \vec{b}))_{\vec{b} \in \mathbb{F}^{H_{r_0}}}\right).$$

A routine computation gives the identities

$$G_{f,r_0,r}(x) = \mathbb{E}_{\vec{a} \in \mathbb{F}^{H_r}} T_{\vec{a} \cdot \vec{v}} g(x)$$

and

$$G_{f,r_0}(x) = \mathbb{E}_{h \in V} T_h g(x) = \int_V g.$$

Also, it is clear that G is Lipschitz with norm $O_{F,r_0}(1)$. The claim then follows from (3.2). \square

Lemma 3.12 (Global polynomiality test can be approximated by local polynomiality test). *Let $k \geq 1$, let V be a finite-dimensional vector space, let $f : V \rightarrow \mathcal{D}$ be a bounded function, and let $v_1, v_2, \dots \in V$ be an accurate sampling sequence for f of degree k and at scales H_1, H_2, \dots . Then for every finite sequence $\vec{b}_1, \dots, \vec{b}_m \in \mathbb{F}^\omega$ and every continuous function $F : \mathcal{D}^m \rightarrow \mathbb{C}$, we have*

$$\begin{aligned} & \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{H_{r_1}}} \dots \mathbb{E}_{\vec{a}_k \in \mathbb{F}^{H_{r_k}}} \int_V |\Delta_{\vec{a}_1 \cdot \vec{v}} \dots \Delta_{\vec{a}_k \cdot \vec{v}} g - 1| \\ &= \mathbb{E}_{h_1, \dots, h_k \in V} \int_V |\Delta_{h_1} \dots \Delta_{h_k} g - 1| + o_{r_1 \rightarrow 0; F, m, \vec{b}_1, \dots, \vec{b}_m, k}(1) \end{aligned}$$

for any $1 \leq r_1 < r_2 < \dots < r_k$, where $g : V \rightarrow \mathbb{C}$ is the function defined by (3.3).

Proof. Arguing as in Lemma 3.11, we may assume that $\{\vec{b}_1, \dots, \vec{b}_m\} = \mathbb{F}^{H_{r_0}}$ for some $r_0 > 0$ depending on $\vec{b}_1, \dots, \vec{b}_m$, and that $F : \mathcal{D}^{\mathbb{F}^{H_{r_0}}} \rightarrow \mathbb{C}$ is Lipschitz.

Note that to prove the claim we may without loss of generality restrict to the regime $r_1 > r_0$. We now apply (3.2) with $G : \mathcal{D}^{\{0,1\}^k \times \mathbb{F}^{H_{r_0}}} \rightarrow \mathbb{C}$ being the function

$$G\left((z(\omega, \vec{b}))_{\omega \in \{0,1\}^k, \vec{b} \in \mathbb{F}^{H_{r_0}}}\right) := \left| \prod_{\omega \in \{0,1\}^k} \mathcal{C}^{\omega_1 + \dots + \omega_k} F\left((z(\omega, \vec{b}))_{\vec{b} \in \mathbb{F}^{H_{r_0}}}\right) - 1 \right|$$

where \mathcal{C} is again the complex conjugation operator. A routine computation gives the identities

$$G_{f, r_0, r_1, \dots, r_k}(x) = \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{H_{r_1}}} \dots \mathbb{E}_{\vec{a}_k \in \mathbb{F}^{H_{r_k}}} |\Delta_{\vec{a}_1 \cdot \vec{v}} \dots \Delta_{\vec{a}_k \cdot \vec{v}} g(x) - 1|$$

and

$$G_{f, r_0}(x) = \mathbb{E}_{h_1, \dots, h_k \in V} |\Delta_{h_1} \dots \Delta_{h_k} g(x) - 1|$$

for any $r_0 < r_1 < \dots < r_k$. Also it is clear that G is Lipschitz with norm $O_{F, r_0, k}(1)$. The claim then follows from (3.2) and the triangle inequality. \square

Of course, in order to utilise the above lemmas we need to know that such accurate sampling sequences in fact exist. This is the purpose of the following proposition.

Proposition 3.13 (Existence of accurate sampling sequence). *Let $d \geq 1$. Then there exists a sequence*

$$0 = H_0 < H_1 < H_2 < H_3 < \dots$$

of integers such that for every finite-dimensional vector space V and any function $f : V \rightarrow \mathcal{D}$, there exists an accurate sampling sequence $v_1, v_2, v_3, \dots \in V$ for f of degree d at scales H_1, H_2, \dots

Remark 3.14. The key point here is that the scales H_1, H_2, H_3, \dots are *universal*; they depend on d , but otherwise and work for all vector spaces V and functions f .

Proof. We select H_j recursively by the formula $H_{j+1} := F(H_j)$, where $F = F_d : \mathbb{N} \rightarrow \mathbb{N}$ is a sufficiently rapidly growing function depending on d that we will choose later.

We use the probabilistic method, choosing $v_1, v_2, \dots \in V$ uniformly at random, and showing that (if F was sufficiently rapid) the resulting sequence will be an accurate sampling sequence with positive probability.

We begin with observing that in order to verify the condition (3.2), it suffices by the triangle inequality to show that with positive probability, one has

$$\int_V |G_{f, r_0, r_1, \dots, r_{d'}} - G_{f, r_0, r_1, \dots, r_{d'-1}}| \leq \frac{\|G\|_{\text{Lip}}}{dr_1} \quad (3.4)$$

for all $1 \leq d' \leq d$, all $0 \leq r_0 < \dots < r_{d'}$, and every Lipschitz function $G : \mathcal{D}^{\{0,1\}^d \times \mathbb{F}^{H_{r_0}}} \rightarrow \mathbb{C}$, where

$$G_{f,r_0,r_1,\dots,r_{d'}}(x) := \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{H_{r_1}}, \dots, \vec{a}_{d'} \in \mathbb{F}^{H_{r_{d'}}}} \mathbb{E}_{h_{d'+1}, \dots, h_d \in V} G \left((f(x + \sum_{j=1}^{d'} \omega_j \vec{a}_j \cdot \vec{v}_j + \sum_{j=d'+1}^d \omega_j h_j + b \cdot \vec{v}_0))_{(\omega_1, \dots, \omega_d) \in \{0,1\}^d, b \in \mathbb{F}^{H_{r_0}}} \right).$$

By the union bound, it will suffice to show that for all $1 \leq d' \leq d$ and all $0 \leq r_0 < \dots < r_{d'}$, with probability $1 - o_{H_{r_{d'}} \rightarrow \infty; d, H_{r_0}, \dots, H_{r_{d'-1}}, r_1}(1)$, (3.4) holds for all Lipschitz functions $G : \mathcal{D}^{\{0,1\}^d \times \mathbb{F}^{H_{r_0}}} \rightarrow \mathbb{C}$, since the total failure probability can be made to be less than 1 by choosing F to be sufficiently rapid.

We can normalise G to have Lipschitz norm 1. By the Arzelá-Ascoli theorem, the space of such functions is compact in the uniform topology. In particular, there exists a collection of functions $G : \mathcal{D}^{\{0,1\}^d \times \mathbb{F}^{H_{r_0}}} \rightarrow \mathbb{C}$ of Lipschitz norm 1, S , of size $O_{d, H_{r_0}, r_1}(1)$, such that any other such Lipschitz function lies within $\frac{1}{4dr_1}$ (say) of a function $G \in S$ in the uniform metric. Because of this, we see from the union bound again that it will suffice to show that for all $1 \leq d' \leq d$ and all $0 \leq r_0 < \dots < r_{d'}$, and all functions $G : \mathcal{D}^{\{0,1\}^d \times \mathbb{F}^{H_{r_0}}} \rightarrow \mathbb{C}$ of Lipschitz norm 1 in S ,

$$\int_V |G_{f,r_0,r_1,\dots,r_{d'}} - G_{f,r_0,r_1,\dots,r_{d'-1}}| \leq \frac{1}{2dr_1} \quad (3.5)$$

of (3.4) holds with probability $1 - o_{H_{r_{d'}} \rightarrow \infty; d, H_{r_0}, \dots, H_{r_{d'-1}}, r_1}(1)$.

Fix $d', r_0, \dots, r_{d'}, G$. By Markov's inequality, it suffices to show that

$$\mathbb{E} \int_V |G_{f,r_0,r_1,\dots,r_{d'}} - G_{f,r_0,r_1,\dots,r_{d'-1}}| = o_{H_{r_{d'}} \rightarrow \infty; d, H_{r_0}, \dots, H_{r_{d'-1}}}(1);$$

by linearity of expectation it thus suffices to show that

$$\mathbb{E} |G_{f,r_0,r_1,\dots,r_{d'}}(x) - G_{f,r_0,r_1,\dots,r_{d'-1}}(x)| = o_{H_{r_{d'}} \rightarrow \infty; d, H_{r_0}, \dots, H_{r_{d'-1}}}(1)$$

uniformly in $x \in V$.

Fix x . We observe that

$$G_{f,r_0,r_1,\dots,r_{d'}}(x) = \mathbb{E}_{\vec{a} \in \mathbb{F}^{H_{r_{d'}}}} f_{v_1, \dots, v_{H_{r_{d'-1}}}}(\vec{a} \cdot \vec{v}_{d'})$$

and

$$G_{f,r_0,r_1,\dots,r_{d'-1}}(x) = \mathbb{E}_{h \in V} f_{v_1, \dots, v_{H_{r_{d'-1}}}}(h)$$

where $f_{v_1, \dots, v_{H_{r_{d'-1}}}} : V \rightarrow \mathbb{C}$ is the function

$$f_{v_1, \dots, v_{H_{r_{d'-1}}}}(h) := \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{H_{r_1}}, \dots, \vec{a}_{d'-1} \in \mathbb{F}^{H_{r_{d'-1}}}} \mathbb{E}_{h_{d'+1}, \dots, h_d \in V} G \left((f(x + \sum_{j=1}^{d'-1} \omega_j \vec{a}_j \cdot \vec{v}_j + \omega_{d'} h_{d'} + \sum_{j=d'+1}^d \omega_j h_j + b \cdot \vec{v}_0))_{(\omega_1, \dots, \omega_d) \in \{0,1\}^d, b \in \mathbb{F}^{H_{r_0}}} \right).$$

As the notation suggests, the function $f_{v_1, \dots, v_{H_{r_{d'-1}}}}$ depends on the values of $v_1, \dots, v_{H_{r_{d'-1}}}$ but not on higher elements of the sequence. Also, as G has Lipschitz norm 1, f takes values in \mathcal{D} . The claim now follows from Corollary 3.3. \square

4. PROOF OF MAIN THEOREMS

We are now ready to prove the main theorems. We shall just prove Theorem 1.10 using Theorem 1.20; the deduction of Theorem 1.9 using Theorem 1.19 is exactly analogous (see the brief remarks at the end of this section).

Fix \mathbb{F} and d , and let $k = C(d)$ be the quantity in Theorem 1.20. By increasing k if necessary we may assume $k \geq d$. Assume for sake of contradiction that Theorem 1.10 failed for this choice of \mathbb{F}, d, k . Then we can find $\delta > 0$ and a sequence $f^{(n)} : V^{(n)} \rightarrow \mathcal{D}$ of functions on finite-dimensional vector spaces $V^{(n)}$ such that

$$\|f^{(n)}\|_{U^d(V^{(n)})} \geq \delta \quad (4.1)$$

for all n , but

$$\|f^{(n)}\|_{u^k(V^{(n)})} = o_{n \rightarrow \infty}(1). \quad (4.2)$$

We now let $F(x) := x$, and let

$$1 < H_1 < H_2 < \dots$$

be the sequence in Proposition 3.13; it is important to note that this sequence does not depend on n . From that proposition, we can find an accurate sampling sequence

$$v_1^{(n)}, v_2^{(n)}, \dots \in V^{(n)}$$

for $f^{(n)}$ of degree k at these scales. We fix such a sequence for each n .

We will use these sampling sequences to lift the functions $f^{(n)}$ on $V^{(n)}$ to a universal dynamical system for \mathbb{F}^ω by the usual Furstenberg correspondence principle method. We begin by constructing this universal space.

Definition 4.1 (Furstenberg universal space). Let $X := \mathcal{D}^{\mathbb{F}^\omega}$ be the space of functions $\zeta : \mathbb{F}^\omega \rightarrow \mathcal{D}$. With the product topology, this is a compact metrisable space, with Borel σ -algebra \mathcal{B} . It has a continuous action $h \mapsto T_h$ of the additive group \mathbb{F}^ω , defined by the formula

$$T_h \zeta(x) := \zeta(x + h).$$

We let $\text{Pr}(X)^T$ be the space of all Borel probability measures μ on X which are invariant with respect to this action; note that $\mathbf{X} = (X, \mathcal{B}, \mu, (T_h)_{h \in \mathbb{F}^\omega})$ is a \mathbb{F}^ω -system for any $\mu \in \text{Pr}(X)^T$. If $\mu^{(n)} \in \text{Pr}(X)^T$ is a sequence of such measures, and $\mu \in \text{Pr}(X)^T$ is another measure, we say that $\mu^{(n)}$ *converges vaguely* to μ if we have

$$\lim_{n \rightarrow \infty} \int_X \phi(\zeta) d\mu^{(n)}(\zeta) \rightarrow \int_X \phi(\zeta) d\mu(\zeta)$$

for all continuous functions $\phi : X \rightarrow \mathbb{C}$.

Because X is compact metrisable, and the action of T is continuous it is a well-known fact that $\text{Pr}(X)^T$ is sequentially compact; thus every sequence of measures in $\text{Pr}(X)^T$ has a vaguely convergent subsequence whose limit is also in $\text{Pr}(X)^T$.

For each n , we define a measure $\mu^{(n)} \in \text{Pr}(X)^T$ on X by the formula

$$\mu^{(n)} = \mathbb{E}_{x \in V^{(n)}} \delta_{\zeta_{n,x}}$$

where δ denotes the Dirac mass and for each $x \in V^{(n)}$, $\zeta_{n,x} \in X$ is the function

$$\zeta_{n,x}(\vec{a}) := T_{\vec{a} \cdot \vec{v}^{(n)}} f^{(n)}(x) = T_{\sum_{m=1}^{\infty} a_m v_m^{(n)}} f^{(n)}(x)$$

for all $\vec{a} \in \mathbb{F}^\omega$ (note the sum on the right-hand side has only finitely many non-zero terms). Observe that $\mu^{(n)}$ is indeed T -invariant. By passing to a subsequence if necessary, we may assume that $\mu^{(n)}$ converges vaguely to a limit $\mu \in \text{Pr}(X)^T$. We write $\mathbf{X} := (X, \mathcal{B}, \mu, (T_h)_{h \in \mathbb{F}^\omega})$.

Let $f : X \rightarrow \mathcal{D}$ be the indicator function $f(\zeta) := \zeta(0)$. We observe the key correspondence

$$\int_X G(T_{\vec{a}_1} f, \dots, T_{\vec{a}_k} f) d\mu^{(n)}(\zeta) = \int_{V^{(n)}} G(T_{\vec{a}_1 \cdot \vec{v}^{(n)}} f^{(n)}, \dots, T_{\vec{a}_k \cdot \vec{v}^{(n)}} f^{(n)}) \quad (4.3)$$

for all $\vec{a}_1, \dots, \vec{a}_k \in \mathbb{F}^\omega$, all n , and all continuous $G : \mathcal{D}^k \rightarrow \mathbb{C}$.

We now record the (standard) fact that the countable collection of shifts $T_h f$ for $h \in \mathbb{F}^\omega$ generate $L^\infty(\mathbf{X})$:

Lemma 4.2 ($T_h f$ generate $L^\infty(\mathbf{X})$). *Given any $\phi \in L^\infty(\mathbf{X})$ and $\varepsilon > 0$, there exists a finite number of shifts $\vec{h}_1, \dots, \vec{h}_k \in \mathbb{F}^\omega$ and a continuous function $G : \mathcal{D}^k \rightarrow \mathbb{C}$ such that*

$$\int_X |\phi - G(T_{\vec{h}_1} f, \dots, T_{\vec{h}_k} f)| d\mu \leq \varepsilon.$$

Proof. For continuous ϕ , the claim follows easily from the Stone-Weierstrass theorem (and in this case we can upgrade the L^1 approximation to L^∞ approximation). As X is compact metrisable, the Borel measure μ is in fact a Radon measure, and so (by Urysohn's lemma) the continuous functions are dense in $L^\infty(\mathbf{X})$ in the $L^1(\mathbf{X})$ topology, and the claim follows. \square

We can now use the machinery of the previous section to deduce various important facts about \mathbf{X} and f . For instance, Lemma 3.11 now implies

Lemma 4.3 (Ergodicity). *\mathbf{X} is ergodic.*

Proof. By the mean ergodic theorem, it suffices to show that

$$\lim_{r \rightarrow \infty} \int_X \left| \mathbb{E}_{\vec{h} \in \mathbb{F}^{H_r}} T_{\vec{h}} g - \int_X g d\mu \right| d\mu = 0$$

for all $g \in L^\infty(X)$. By Lemma 4.2 and a standard limiting argument it suffices to show this for g which are functions of finitely many shifts of f , say $g = G(T_{\vec{b}_1} f, \dots, T_{\vec{b}_k} f)$. We will then show that

$$\int_X \left| \mathbb{E}_{\vec{h} \in \mathbb{F}^{H_r}} T_{\vec{h}} g - \int_X g d\mu \right| d\mu = o_{r \rightarrow \infty; G, k, \vec{b}_1, \dots, \vec{b}_k}(1).$$

By vague convergence it suffices to show that

$$\int_X \left| \mathbb{E}_{\vec{h} \in \mathbb{F}^{H_r}} T_{\vec{h}} g - \int_X g d\mu^{(n)} \right| d\mu^{(n)} = o_{r \rightarrow \infty; G, k, \vec{b}_1, \dots, \vec{b}_k}(1)$$

for all n . By (4.3), we can rewrite the left-hand side as

$$\int_V \left| \mathbb{E}_{\vec{h} \in \mathbb{F}^{H_r}} T_{\vec{h} \cdot \vec{v}_r^{(n)}} g^{(n)} - \int_V g^{(n)} \right|$$

where

$$g^{(n)} := G(T_{\vec{b}_1 \cdot \vec{v}^{(n)}} f^{(n)}, \dots, T_{\vec{b}_k \cdot \vec{v}^{(n)}} f^{(n)}).$$

But the claim now follows from Lemma 3.11 (and Remark 3.8). \square

In a similar spirit, Lemma 3.9 implies

Lemma 4.4 (f has large Gowers-Host-Kra norm). *We have $\|f\|_{U^d(\mathbf{X})} \geq \delta$.*

Proof. From the mean ergodic theorem we have

$$\|f\|_{U^1(\mathbf{X})}^2 = \limsup_{K_1 \rightarrow \infty} \mathbb{E}_{\vec{h}_1 \in \mathbb{F}^{K_1}} \int_X \Delta_{\vec{h}_1} f d\mu$$

and by induction we have

$$\|f\|_{U^d(\mathbf{X})}^{2^d} = \limsup_{K_d \rightarrow \infty} \dots \limsup_{K_1 \rightarrow \infty} \mathbb{E}_{\vec{h}_d \in \mathbb{F}^{K_d}} \dots \mathbb{E}_{\vec{h}_1 \in \mathbb{F}^{K_1}} \int_X \Delta_{\vec{h}_1} \dots \Delta_{\vec{h}_d} f d\mu.$$

It thus suffices to show that

$$\mathbb{E}_{\vec{h}_d \in \mathbb{F}^{H_{r_d}}} \dots \mathbb{E}_{\vec{h}_1 \in \mathbb{F}^{H_{r_1}}} \int_X \Delta_{\vec{h}_1} \dots \Delta_{\vec{h}_d} f d\mu > \delta^{2^d} - o_{r_d \rightarrow \infty}(1)$$

whenever

$$1 \leq r_d < \dots < r_1.$$

By reversing the order of averages, it suffices to show that

$$\mathbb{E}_{\vec{h}_d \in \mathbb{F}^{H_{r_d}}} \dots \mathbb{E}_{\vec{h}_1 \in \mathbb{F}^{H_{r_1}}} \int_X \Delta_{\vec{h}_1} \dots \Delta_{\vec{h}_d} f d\mu > \delta^{2^d} - o_{r_1 \rightarrow \infty}(1)$$

whenever

$$1 \leq r_1 < \dots < r_d.$$

Fix r_1, \dots, r_d . By weak convergence, it suffices to show that

$$\mathbb{E}_{\vec{h}_d \in \mathbb{F}^{H_{r_d}}} \dots \mathbb{E}_{\vec{h}_1 \in \mathbb{F}^{H_{r_1}}} \int_X \Delta_{\vec{h}_1} \dots \Delta_{\vec{h}_d} f d\mu^{(n)} > \delta^{2^d} - o_{r_1 \rightarrow \infty}(1)$$

for all n . By (4.1), it suffices to show that

$$\mathbb{E}_{\vec{h}_d \in \mathbb{F}^{H_{r_d}}} \dots \mathbb{E}_{\vec{h}_1 \in \mathbb{F}^{H_{r_1}}} \int_X \Delta_{\vec{h}_1} \dots \Delta_{\vec{h}_d} f d\mu^{(n)} > \|f^{(n)}\|_{U^d(V^{(n)})}^{2^d} - o_{r_1 \rightarrow \infty}(1).$$

By (4.3), left-hand side can be rephrased as

$$\int_V \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{H_{r_1}}, \dots, \vec{a}_d \in \mathbb{F}^{H_{r_d}}} \Delta_{\vec{a}_1 \cdot \vec{v}_{r_1}^{(n)}} \dots \Delta_{\vec{a}_d \cdot \vec{v}_{r_d}^{(n)}} f^{(n)}$$

and the claim now follows from Lemma 3.9 (and Remark 3.8). \square

We have now verified all the hypotheses of Theorem 1.19. Applying that theorem, we conclude that $\|f\|_{u^k(\mathbf{X})} > c$ for some $c > 0$ (which could be very small, but positive). Thus we can find a phase polynomial $\phi \in \mathcal{P}_{k-1}(\mathbf{X})$ of degree $k-1$ such that

$$\left| \int_X f \bar{\phi} \, d\mu \right| > c.$$

Let $\varepsilon > 0$ be a small number (depending on d, k, c) to be chosen later. By Lemma 4.2, we can find $\vec{b}_1, \dots, \vec{b}_m \in \mathbb{F}^\omega$ (with m potentially quite large, but finite) and a continuous $G : \mathcal{D}^m \rightarrow \mathbb{C}$ such that

$$\int_X |\phi - G(T_{\vec{b}_1} f, \dots, T_{\vec{b}_m} f)| \leq \varepsilon. \quad (4.4)$$

Since ϕ takes values in \mathcal{D} , we may assume without loss of generality that G does also. If ε is small enough depending on c , we thus have

$$\left| \int_X f \overline{G(T_{\vec{b}_1} f, \dots, T_{\vec{b}_m} f)} \, d\mu \right| > c/2.$$

By vague convergence, we thus have

$$\left| \int_X f \overline{G(T_{\vec{b}_1} f, \dots, T_{\vec{b}_m} f)} \, d\mu^{(n)} \right| > c/4$$

for all sufficiently large n (depending on G, m, c). Using (4.3), we rearrange this as

$$\left| \int_V f^{(n)} \overline{G(T_{\vec{b}_1 \cdot \vec{v}^{(n)}} f^{(n)}, \dots, T_{\vec{b}_m \cdot \vec{v}^{(n)}} f^{(n)})} \right| > c/4. \quad (4.5)$$

Now let r_1 be a large integer depending on the $\vec{b}_1, \dots, \vec{b}_m, \varepsilon$, and let $r_j := r_1 + (j-1)$ for $j = 2, \dots, d$. Since ϕ is a phase polynomial of degree $k-1$, we have

$$\int_X |\Delta_{\vec{a}_1} \dots \Delta_{\vec{a}_k} \phi - 1| \, d\mu = 0$$

for all $\vec{a}_1 \in \mathbb{F}^{H_{r_1}}, \dots, \vec{a}_k \in \mathbb{F}^{H_{r_k}}$. From many applications of (4.4), the triangle inequality, and the boundedness of ϕ, G , we conclude that

$$\int_X |\Delta_{\vec{a}_1} \dots \Delta_{\vec{a}_k} G(T_{\vec{b}_1} f, \dots, T_{\vec{b}_m} f) - 1| \, d\mu \ll_k \varepsilon$$

for all $\vec{a}_1 \in \mathbb{F}^{H_{r_1}}, \dots, \vec{a}_k \in \mathbb{F}^{H_{r_k}}$. By vague convergence, this implies that

$$\int_X |\Delta_{\vec{a}_1} \dots \Delta_{\vec{a}_k} G(T_{\vec{b}_1} f, \dots, T_{\vec{b}_m} f) - 1| \, d\mu^{(n)} \ll_k \varepsilon$$

for all sufficiently large n (depending on $\varepsilon, H_{r_1}, \dots, H_{r_k}$). Using (4.3), we can rearrange the left-hand side as

$$\int_{V^{(n)}} |\Delta_{\vec{a}_1 \cdot \vec{v}^{(n)}} \dots \Delta_{\vec{a}_k \cdot \vec{v}^{(n)}} G(T_{\vec{b}_1 \cdot \vec{v}^{(n)}} f^{(n)}, \dots, T_{\vec{b}_m \cdot \vec{v}^{(n)}} f^{(n)}) - 1|$$

and so on averaging we obtain

$$\mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{Hr_1}, \dots, \vec{a}_k \in \mathbb{F}^{Hr_k}} \int_{V^{(n)}} |\Delta_{\vec{a}_1 \cdot \vec{v}^{(n)}} \dots \Delta_{\vec{a}_k \cdot \vec{v}^{(n)}} G(T_{\vec{b}_1 \cdot \vec{v}^{(n)}} f^{(n)}, \dots, T_{\vec{b}_m \cdot \vec{v}^{(n)}} f^{(n)}) - 1| \ll_k \varepsilon.$$

Applying Lemma 3.12 we conclude (if r_1 is sufficiently large depending on $\vec{b}_1, \dots, \vec{b}_m, \varepsilon$) that

$$\mathbb{E}_{h_1, \dots, h_k \in V^{(n)}} \int_{V^{(n)}} |\Delta_{h_1} \dots \Delta_{h_k} G(T_{\vec{b}_1 \cdot \vec{v}^{(n)}} f^{(n)}, \dots, T_{\vec{b}_m \cdot \vec{v}^{(n)}} f^{(n)}) - 1| \ll_k \varepsilon.$$

Now we invoke a local testability lemma:

Lemma 4.5 (Polynomiality is locally testable). *Let V be a finite-dimensional vector space, let $k \geq 1$, let $g : V \rightarrow \mathcal{D}$ be a bounded function, and suppose that*

$$\mathbb{E}_{h_1, \dots, h_k \in V} \int_V |\Delta_{h_1} \dots \Delta_{h_k} g - 1| \leq \varepsilon \tag{4.6}$$

for some $\varepsilon > 0$. Then there exists a phase polynomial $\phi \in \mathcal{P}_{k-1}(V)$ such that

$$\int_V |g - \phi| \leq o_{\varepsilon \rightarrow 0; d}(1).$$

For $\mathbb{F} = \mathbb{F}_2$, this result is essentially in [1] or [20, Proposition 4.6], but for the convenience of the reader (and in view of the subtle difference between phase polynomials and polynomials, see Remark 1.3) we give a full proof of this lemma in the appendix.

Applying this lemma, we conclude that there exists $\phi^{(n)} \in \mathcal{P}_{k-1}(V^{(n)})$ such that

$$\int_V |G(T_{\vec{b}_1 \cdot \vec{v}^{(n)}} f^{(n)}, \dots, T_{\vec{b}_m \cdot \vec{v}^{(n)}} f^{(n)}) - \phi^{(n)}| \leq o_{\varepsilon \rightarrow 0; k}(1).$$

Inserting this into (4.5) we conclude that

$$|\int_V f^{(n)} \overline{\phi^{(n)}}| > c/8$$

if ε is sufficiently small depending on c, k . But this contradicts (4.2). The proof of Theorem 1.10 is complete.

The proof of Theorem 1.9 is identical, but with k now set equal to d , and Theorem 1.19 used instead of Theorem 1.20. We leave the details to the reader.

Remark 4.6. It is tempting to try to adapt these arguments to the cyclic setting $\mathbb{Z}/N\mathbb{Z}$, in which the role of polynomials is replaced by that of a *nilsequence* (see [10], [11] for further discussion), thus establishing the *Inverse conjecture for the Gowers norm* for $\mathbb{Z}/N\mathbb{Z}$ that was formulated in those papers. The analogue of Theorem 1.19 is known, see [14]. However, two obstructions remain before one can carry out this program. The first is to compensate for the rigidity of arithmetic progressions that seems to prevent a counterpart of Corollary 3.3 from holding in the cyclic group setting (see Remark 3.4). The second is that whereas polynomiality is locally testable thanks to Lemma 4.5, it is unclear whether the property of being a nilsequence is similarly testable.

APPENDIX A. PROOF OF LEMMA 4.5

In this appendix we give a proof of Lemma 4.5, following the arguments in [1] and [20, Proposition 4.6]. We begin with a variant of Lemma 1.2:

Lemma A.1 (Discreteness). *Let $k \geq 0$, let V be a finite-dimensional vector space, and $\phi \in \mathcal{P}_k(V)$. Then there exists $\theta \in \mathbb{R}/\mathbb{Z}$ and an integer $K \geq 1$ depending only on \mathbb{F} such that $\phi(x)$ is equal to $e^{2\pi i \theta}$ times a K^{th} root of unity for every $x \in V$.*

Proof. See [2, Lemma D.5] (which gives the explicit value $K = p^{\lfloor k/p \rfloor + 1}$, where p is the characteristic of \mathbb{F}). \square

We also have a rigidity lemma.

Lemma A.2 (Rigidity). *Let $k \geq 0$, let V be a finite-dimensional vector space, and $\phi \in \mathcal{P}_k(V)$. Suppose that $\int_V |\phi - 1| \leq \varepsilon$ for some $\varepsilon > 0$. If ε is sufficiently small depending on k, \mathbb{F} , then ϕ is constant.*

Proof. We induct on k . For $k = 0$ the claim is obvious, and for $k = 1$ ϕ is a linear character (times a phase) and the claim can be worked out by hand. Now suppose $k \geq 2$ and the claim has already been shown for smaller values of k . Since ϕ is a phase polynomial, we have $\Delta_0 \dots \Delta_0 \phi = 1$, and thus ϕ has unit magnitude. Observe that if $\int_V |\phi - 1| \leq \varepsilon$, then $\int_V |T_h \phi - 1| \leq \varepsilon$ for every $h \in V$. Using the elementary estimate

$$|\Delta_h \phi - 1| \leq |\phi - 1| + |T_h \phi - 1|$$

(using the fact that ϕ has unit magnitude) we conclude that

$$\int_V |\Delta_h \phi - 1| \leq 2\varepsilon$$

for every $h \in V$. On the other hand, $\Delta_h \phi \in \mathcal{P}_{k-1}(V)$, so by induction hypothesis (if ε is small enough) we conclude that $\Delta_h \phi$ is constant for all $h \in V$. Thus $\phi \in \mathcal{P}_1(V)$, but then the claim follows from the $k - 1$ case. \square

We now prove Lemma 4.5. The case $k = 1$ is easy, so suppose that $k \geq 2$ and the claim has already been established for $k - 1$. To abbreviate the notation we shall write $o(1)$ for $o_{\varepsilon \rightarrow 0; k}(1)$. We say that a statement $P(x)$ holds for *most* $x \in V$ if it holds for $(1 - o(1))|V|$ elements of v .

We fix k, V, f . We may assume that ε is small depending on d , as the claim is trivial otherwise. From (4.6) and Markov's inequality we see that

$$\mathbb{E}_{h_1, \dots, h_{k-1} \in V} \int_V |\Delta_{h_1} \dots \Delta_{h_{k-1}} \Delta_h f - 1| = o(1) \tag{A.1}$$

for most $h \in V$. Let us call h *good* if (A.1) holds. Applying the induction hypothesis, we conclude that for any good h there exists² $\phi_h \in \mathcal{P}_{k-2}(V)$ such that

$$\int_V |\Delta_h f - \phi_h| \leq o(1).$$

In particular, this implies (by Markov's inequality) that for all good h , we have

$$f(x+h)\overline{f(x)} = \phi_h(x) + o(1)$$

for most V . Since f is bounded in magnitude by 1, this implies that

$$|f(x)| = 1 - o(1)$$

for most x , and for all good h we have

$$f(x+h) = \phi_h(x)f(x) + o(1) \tag{A.2}$$

for most x .

We now pause to perform a discretisation trick. Write $p := \text{char}(\mathbb{F})$. From repeated applications of (A.2) we see that

$$f(x) = f(x+ph) = \phi_h(x)\phi_h(x+h) \dots \phi_h(x+(p-1)h)f(x) + o(1)$$

for most x , and thus

$$\phi_h(x)\phi_h(x+h) \dots \phi_h(x+(p-1)h) = 1 + o(1)$$

for at least one x . On the other hand, from Lemma A.1 ϕ_h takes values in $e^{2\pi i\theta}$ times K^{th} roots of unity for some fixed K depending only on d, p . Thus $e^{2\pi i p\theta}$ times a K^{th} root of unity is within $o(1)$ of 1, and so $e^{2\pi i\theta}$ lies within $o(1)$ of a pK^{th} root of unity. Rotating ϕ_h by $o(1)$ if necessary we may assume that $e^{2\pi i\theta}$ is *exactly* a pK^{th} root of unity, and in particular we have

$$\phi_h^{pK} \equiv 1 \tag{A.3}$$

whenever h is good.

Now suppose that h_1, h_2, h_3, h_4 are good and form an *additive quadruple* in the sense that $h_1 + h_2 = h_3 + h_4$. Then from (A.2) we see that

$$f(x+h_1+h_2) = f(x)\phi_{h_1}(x)\phi_{h_2}(x+h_1) + o(1) \tag{A.4}$$

for most x , and similarly

$$f(x+h_3+h_4) = f(x)\phi_{h_3}(x)\phi_{h_4}(x+h_3) + o(1)$$

for most x . Since $|f(x)| = 1 + o(1)$ for most x , we conclude the approximate cocycle relationship

$$\phi_{h_1}(x)\phi_{h_2}(x+h_1)\overline{\phi_{h_3}(x)\phi_{h_4}(x+h_3)} = 1 + o(1)$$

for most x . In particular, the average of the left-hand side in x is $1 - o(1)$. Applying Lemma A.2 (and assuming ε small enough), we conclude that the left-hand side is *constant* in x ; using the discretisation (A.3), we conclude (again for ε small enough) that it is in fact 1. Thus

$$\phi_{h_1}(x)\phi_{h_2}(x+h_1) = \phi_{h_3}(x)\phi_{h_4}(x+h_3) \tag{A.5}$$

for *all* x and any good additive quadruple h_1, h_2, h_3, h_4 .

²This quantity plays the same role that *cocycles* do in ergodic theory.

Now for any $k \in V$, define the quantity $\psi(k) \in \mathbb{C}$ by the formula

$$\psi(k) := \phi_{h_1}(0)\phi_{h_2}(h_1) \quad (\text{A.6})$$

whenever $h_1, h_2, h_1 + h_2$ are simultaneously good. Note that the existence of such an h_1, h_2 is guaranteed since most h are good, and (A.5) ensures that the right-hand side of (A.6) does not depend on the exact choice of h_1, h_2 and so ψ is well-defined. From (A.3) we see that ψ takes values in the pK^{th} roots of unity, and in particular only has $O(1)$ possible values.

Now let $x \in V$ and h be good. Then, since most elements of V are good, we can find good r_1, r_2, s_1, s_2 such that $r_1 + r_2 = x$ and $s_1 + s_2 = x + h$. From (A.4) we see that

$$f(y + x) = f(y + r_1 + r_2) = f(y)\phi_{r_1}(y)\phi_{r_2}(y + r_1) + o(1)$$

and

$$f(y + x + h) = f(y + s_1 + s_2) = f(y)\phi_{s_1}(y)\phi_{s_2}(y + s_1) + o(1)$$

for most y . Also from (A.2) we have

$$f(y + x + h) = f(y + x)\phi_h(y + x) + o(1)$$

for most y . Combining these (and the fact that $|f(y)| = 1 + o(1)$ for most y) we see that

$$\phi_{s_1}(y)\phi_{s_2}(y + s_1)\overline{\phi_{r_1}(y)\phi_{r_2}(y + r_1)\phi_h(y + x)} = 1 + o(1)$$

for most y . Taking expectations and applying Lemma A.2 and (A.3) as before, we conclude that

$$\phi_{s_1}(y)\phi_{s_2}(y + s_1)\overline{\phi_{r_1}(y)\phi_{r_2}(y + r_1)\phi_h(y + x)} = 1$$

for *all* y . Specialising to $y = 0$ and applying (A.6) we conclude that

$$\phi_h(x) = \psi(x + h)\overline{\psi(x)} = \Delta_h \psi(x) \quad (\text{A.7})$$

for all $x \in V$ and good h ; thus we have successfully “integrated” ϕ_h . We can then extend $\phi_h(x)$ to all $h \in V$ (not just good h) by viewing (A.7) as a *definition*. Observe that if $h \in V$, then $h = h_1 + h_2$ for some good h_1, h_2 , and from (A.7) we have

$$\phi_h(x) = \phi_{h_1}(x)\phi_{h_2}(x + h_1).$$

In particular, since the right-hand side lies in $\mathcal{P}_{k-2}(V)$, the left-hand side does also. Thus we see that $\Delta_h \psi \in \mathcal{P}_{k-2}(V)$ for all $h \in V$, and thus $Q \in \mathcal{P}_{k-1}(V)$. If we then set $g(x) := f(x)\overline{\psi(x)}$, then from (A.2), (A.7) we see that for every $h \in H$ we have

$$g(x + h) = g(x) + o(1)$$

for most x . From Fubini’s theorem, we thus conclude that there exists an x such that $g(x + h) = g(x) + o(1)$ for most h , thus g is almost constant. Since $|g(x)| = 1 + o(1)$ for most x , we thus conclude the existence of a phase $\theta \in \mathbb{R}/\mathbb{Z}$ such that $g(x) = e^{2\pi i \theta} + o(1)$ for most x . We conclude that

$$f(x) = e^{2\pi i \theta} \psi(x) + o(1)$$

for most x , and Lemma 4.5 then follows.

REFERENCES

- [1] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn and D. Ron, *Testing low-degree polynomials over $GF(2)$* , RANDOM-APPROX 2003, 188–199. Also: Testing Reed-Muller codes, IEEE Transactions on Information Theory 51 (2005), 4032–4039.
- [2] V. Bergelson, T. Tao and T. Ziegler, *An inverse theorem for the uniformity seminorms associated with the action of \mathbb{F}_p^∞* , preprint.
- [3] A. Bogdanov, E. Viola, *Pseudorandom bits for polynomials*, Proc. of FOCS 2007, 41–51.
- [4] N. Frantzikinakis, B. Host, B. Kra, *Multiple recurrence and convergence for sequences related to the prime numbers*, preprint.
- [5] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.
- [6] T. Gowers, *A new proof of Szemerédi’s theorem for arithmetic progressions of length four*, Geom. Func. Anal. **8** (1998), 529–551.
- [7] T. Gowers, *A new proof of Szemerédi’s theorem*, Geom. Func. Anal., **11** (2001), 465–588.
- [8] T. Gowers, J. Wolf, *The true complexity of a system of linear equations*, preprint.
- [9] B. Green, T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Annals of Math., Annals of Math. **167** (2008), 481–547.
- [10] B. Green, T. Tao, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proc. Edin. Math. Soc. **51** (2008), 73–153.
- [11] B. Green, T. Tao, *Linear equations in primes*, Annals of Math., to appear.
- [12] B. Green, T. Tao, *New bounds for Szemerédi’s Theorem, I: Progressions of length 4 in finite field geometries*, Proc. Lond. Math. Soc. **98** (2009), 365–392.
- [13] B. Green, T. Tao, *The distribution of polynomials over finite fields, with applications to the Gowers norms*, preprint.
- [14] B. Host and B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488.
- [15] S. Lovett, R. Meshulam, A. Samorodnitsky, *Inverse conjecture for the Gowers norm is false*, STOC 2008.
- [16] T. Kaufman, S. Lovett, *Worst Case to Average Case Reductions for Polynomials*, FOCS 2008: 166–175.
- [17] A. Samorodnitsky, *Low-degree tests at large distances*, STOC 2007.
- [18] M. Sudan, L. Trevisan, S. Vadhan, *Pseudorandom generators without the XOR lemma*, Special issue on the Fourteenth Annual IEEE Conference on Computational Complexity (Atlanta, GA, 1999).
- [19] T. Tao, *A quantitative ergodic theory proof of Szemerédi’s theorem*, Electron. J. Combin. **13** (2006) 1 No. 99, 1–49.
- [20] T. Tao, *Structure and randomness in combinatorics*, Proceedings of the 48th annual symposium on Foundations of Computer Science (FOCS) 2007, 3–18.
- [21] T. Tao and V. Vu, Additive Combinatorics, Cambridge Univ. Press, 2006.
- [22] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. **34** (1959) 358–360.

UCLA DEPARTMENT OF MATHEMATICS, LOS ANGELES, CA 90095-1596.

E-mail address: tao@math.ucla.edu

DEPARTMENT OF MATHEMATICS, TECHNION, HAIFA, ISRAEL 32000

E-mail address: tamarzr@tx.technion.ac.il